学术社交网络问答质量智能评价与服务优化研究*

■ 严炜炜　黄为　温馨
武汉大学信息管理学院　武汉 430072

摘　要：［目的/意义］学术社交网络所提供的问答服务已成为学者们快速获取学术信息、解决学术问题的重要途径，实现基于机器学习的问答质量智能评价和服务优化对学术社交网络中优质内容传播具有重要意义。［方法/过程］以ResearchGate问答服务为研究对象，从结构化特征、内容特征、其他特征以及回答者特征4个维度构建答案质量评价体系，利用机器学习方法和数据增强技术进行答案质量分类预测。［结果/结论］SMOTE算法在处理不平衡样本时具备有效性；支持向量机在单一模型预测中，取得出色的分类效果；组合模型使预测精度得到进一步提升，基于随机森林、支持向量机、BP神经网络构建的组合模型分类性能最佳，以此为基础可通过搭建问答质量智能评价系统实现学术社交网络问答服务优化。

关键词：答案质量评价　问答服务　学术社交网络　机器学习
分类号：G203
DOI：10.13266/j.issn.0252-3116.2021.06.014

1　引言

社会化媒体的普及和繁荣促使越来越多的学者利用社会化媒体获取和分享学术信息[1]，因此也催生了专门针对学者设计的学术社交网络平台，如Academia、ResearchGate（RG）等。为了满足学者们多元化的需求、丰富学术信息的传播途径，学术社交网络平台纷纷设计并提供了独具特色的功能与服务以进一步完善学术社交服务体系。其中，RG提供的学术问答服务已经成为学者们快速获取学术信息、增进彼此间交流的重要途径，其由用户自主提出个性化问题，其他具有相关领域知识或经验的用户负责解决问题并提供个性化的答案[2]。

尽管学术社交网络平台问答服务为用户提供了更具针对性的信息获取与交流方式，由于学术社交网络秉持自由开放精神，任何用户都可以随意地提问和回答问题，造成了答案质量控制存在障碍[3]，虚假信息、过时老化信息、冗余信息等突出的信息质量问题给用户带来困扰[4]。此外，用户虽可通过点赞、评论等形式对问答社区中的问答质量提供一定程度的反馈[5]，但马太效应的存在仍使众多高质量答案因发布时间晚、曝光量低而被湮没。

在学术社交网络情境下，问答信息质量对用户满意具有决定性作用[3]，其质量的高低代表了对用户期望的满足程度[6]。利用机器学习或深度学习模型在海量的问答交互数据中进行答案质量自动化评价，自动挖掘和发现高质量答案，筛除和过滤低质内容，可以提升学术问答服务信息质量水平，满足用户对信息质量的期望，是学术社交网络平台提高竞争优势、实现良性发展的重要前提[7]。因此，笔者聚焦代表性学术社交网络平台的问答交互数据，在构建能够客观反映答案质量的评价体系的基础上，利用机器学习方法和数据增强技术进行答案质量的智能分类预测，并通过搭建问答质量智能评价系统实现学术社交网络问答服务的优化。

2　相关研究

社会化问答服务答案质量控制存在障碍、信息质量问题突出的特点吸引了众多研究者的关注，相关研究主要立足于问答社区，探究答案质量的影响因素和答案质量评价两个方面。

关于问答社区答案质量影响因素的研究，国内外

* 本文系国家自然科学基金青年项目“群体差异视角下学术社交网络用户交互与合作机制研究”（项目编号：71904148）研究成果之一。
作者简介：严炜炜（ORCID：0000-0001-6688-3393），副教授，博士，E-mail：yanww@whu.edu.cn；黄为（ORCID：0000-0001-8409-0813），硕士研究生；温馨（ORCID：0000-0002-0797-9069），硕士研究生。
收稿日期：2020-10-07　修回日期：2021-01-12　本文起止页码：129-137　本文责任编辑：徐健

学者常基于数据质量框架[8]、信息质量评价标准[9]等经典理论探索和验证不同因素对答案质量的影响，如基于数据质量框架和信息质量评价标准，孙晓宁等[10]从内容质量、情境质量、来源质量和情感质量4个维度，实证构建了社会化搜索答案质量评价模型；D. Ishikawa等[11]构建了包括回答者经验、证据来源、礼貌程度、详细程度等12个维度的社会化问答社区答案质量评价指标体系。此外，亦有学者将用户感知、外部线索等角度引入答案质量影响因素研究中以完善答案质量评价标准。如在基于用户感知方面，吴雅威等[12]利用信息构建理论从答案、用户和社区视角出发对学术问答社区用户生成答案质量进行评价，并结合情绪理论、认知理论、用户行为等理论对指标进行阐释；在基于外部线索方面，张煜轩[13]结合线索理论和用户视角，发现信息利用线索、信息认同线索、信息举报线索等7类外部线索对用户感知判断社会化问答社区信息质量产生影响。

关于问答社区答案质量评价的研究，目前学者们一般将答案质量评价视为基于机器学习的分类问题[4]，研究的核心主要在"答案质量影响因素的选取与组合"以及"预测算法的选择与优化"上。在影响因素的选取与组合方面，Y. Z. Cai和孔维泽等[14-15]提出了反映答案动态特性的时序特征，并与其他传统特征进行比较，发现基于时间的特征能更好地预测最佳答案；姜雯等[16-17]认为问答社区中的答案具有一定的情感，分别将附加情感标注的特征和社会情感特征引入到答案质量自动化评价中，结果表明情感特征的加入能够提高分类预测的准确率。而在算法的选择与优化方面，L. Li等[18]对RG平台的回答质量展开研究，发现优化后的支持向量机算法在精度上比其他模型有着压倒性的优势；郭顺利等[16]基于GA-BP神经网络模型设计答案质量自动化评价方法，并选取"知乎"数据进行应用研究，实验分析发现GA-BP神经网络相比于其他算法准确率较高、平均误差低；L. T. Le等[19]利用Brainly上收集的问答数据验证其构建的答案质量评估框架的有效性，实验表明个人和社区特征对答案质量预测更为有效，随机森林模型能够准确识别高质量答案。此外，除了使用经典的机器学习方法，亦有学者使用深度学习模型开展答案质量评价研究，如D. V. Vekariya等[20]利用通用的全局最大池化层压缩参数的数量、减少计算量，并结合DeepLSTM模型来预测既定问题的最佳答案；贺勋等[21]采用BTM主题模型计算"问答对"的语义相似度，利用BERT模型将答案评论情感极性转化为答案质量评价特征，最终使用Rank SVM算法融合多特征预测中文问答社区的答案质量。

通过对相关研究的梳理，不难发现国内外学者在探索问答社区答案质量影响因素时，除了延续数据质量框架等经典理论外，还尝试融合不同理论、立足于新颖视角拓展并完善答案质量评价标准。然而，无论是答案质量影响因素研究亦或是答案质量评价研究，相关研究大多集中在综合性问答社区情境下（如：Yahoo! Answers、Quora、知乎等），虽然产生了较多研究成果，但提出的答案质量评价体系和自动化评价模型在各类垂直领域问答社区中的有效性还有待验证。综合性问答社区上的用户大多是匿名注册且背景复杂，而学术社交网络平台则要求用户在注册时提供真实姓名以及所属大学或研究机构的相关信息，这确保了学术问答服务上的答案都是由从事科研工作的人员提供。差异化的用户群体使得这两个平台上的答案也表现一些不同的特点，学术问答服务的回答通常包含更复杂、更专业的内容[22]，可能会阐述一个学术问题的基本理论、方法、内涵和外延[23]。在开展学术问答服务的答案质量评价时，研究者发现学术问答服务的答案质量评价与综合性问答社区的答案质量评价标准具有明显差异[24]，因而有必要探索学术问答服务评价答案质量的新特征。此外，针对答案质量评价研究，国内外学者的研究重心都集中在"特征和算法"上，少有学者从数据增强角度探究预测精度提升的方法，且相关研究缺乏对学术社交网络平台方的指导，即如何在学术社交网络平台的问答服务中实现答案质量自动化评价，进而优化平台问答服务，都是亟待进一步研究与探讨的问题。

3 研究设计

RG是当今最受欢迎的学术社交网络平台[25]，其自2008年上线至今，平台上发布的出版物、问答、研究项目等已超过1.35亿，用户数量也已经突破1 700万。同时，相较于其他学术社交网络平台，RG在庞大用户群体的基础上借助问答功能积累了丰富的问答交互数据资源，可支撑问答质量智能评价研究。考虑到问答质量评价通常针对回答行为和内容开展评价工作，故本研究以RG为对象，在构建答案质量评价体系和自动化评价模型的基础之上，开展基于问答质量智能评价的学术社交网络问答服务优化实证研究。

相比于人工评价答案质量，机器自动评价具有速度快、精度高、成本低等全方位优势，能够满足学术社交网络平台识别高质量答案、优化问答服务的需求，因

此笔者采用机器学习方法实现 RG 平台的问答质量智能评价。本文研究框架由评价体系构建与模型选择、数据准备、建模预测 3 个部分组成,其中评价体系构建与模型选择包括答案质量评价体系的构建与指标量化和评价模型选择;数据准备包括数据采集、数据清洗(缺失值及异常值的处理等)、数据转换(自动化提取特征)、数据标注(人工标注构建监督学习问题);建模预测包括模型评估与比较、模型优化、使用组合模型实现分类预测等步骤,见图 1。

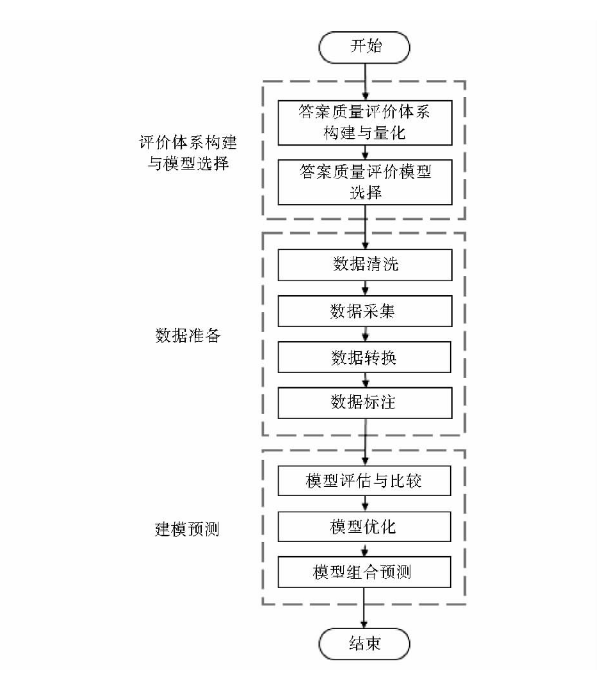


图 1　问答质量智能评价研究框架

3.1　答案质量评价体系

考虑到答案质量评价指标的科学性、有效性以及完整性,笔者在进行广泛文献调研的基础上,选取研究已证实会对答案质量产生影响的评价指标构建出初步的评价体系,包含答案结构化特征、答案内容特征以及答案其他特征 3 个维度指标,并进一步针对 RG 平台的功能特点对回答者特征维度指标进行补充,以构建完整的 RG 答案质量评价体系,见表 1。

表 1　RG 答案质量评价体系

维度	指标	解释及说明	主要参考文献来源
答案结构化特征	文本长度	答案包含的字符数	[16,26]
	关键词数量	答案包含的关键词数量	[16－17]
	句子数量	答案包含的句子数量	[16－17]
	长句占比	答案中长句数量与句子总数的比	[26]
	标点符号占比	答案中标点符号数量与字符数的比	[16,26]
	问题答案长度比	问题长度与答案长度的比	[17]
答案内容特征	问题答案主题相似度	问题与答案的主题相似程度	[26]
	文本多样性	答案的用词多样性	[26]
	答案信息熵	答案包含的信息量	[26]
	答案情感态度	答案所呈现出的回答者情感倾向	[16－17]
	答案主观性	答案的主观程度	[26－27]
答案其他特征	答案回答顺序	答案在所有答案中的位置,按回答时间升序排列	[15,17]
	答案推荐量	答案被赞同/支持的数量	[17,26]
回答者特征	RG-score	RG 制定的一个评价用户的指标	[15,17]
	RI	RG 制定的用于衡量用户对科研成果研究兴趣的指标	[15,17]
	引文量	回答者所有项目或研究资料在 RG 上被引用的总数	[15,17]
	个人获赞量	回答者收获 RG 其他用户推荐的总数	[15,26]
	阅读量	回答者所有项目或研究资料在 RG 上浏览量总和	自设
	项目数	回答者在 RG 上发布的科研项目数量	自设
	回答数	回答者在 RG 上回答问题的数量	[16－17]
	提问数	回答者在 RG 上提出问题的数量	[16－17]
	回答者头像	回答者是否拥有头像	自设
	研究项目数量	回答者在 RG 上的发布或认领的研究资料数量,包括期刊论文、会议论文、预印本等类型	自设

3.1.1 答案结构化特征

结构化特征是指可以直接由答案统计得到的特征。与传统的网页资源相同,学术社交网络平台问答服务的答案主要以文本的形式呈现。因此,适用于分析传统网页质量的相关指标及方法都可以应用于学术问答服务中答案质量的评价,具体包括文本长度、关键词数量、标点符号占比等。答案结构化特征的量化较为简单,借助分词工具和文本处理技术,编写 python 程序可直接从答案文本中进行指标的统计和量化。

3.1.2 答案内容特征

内容特征是指蕴含在文本中,需要通过自然语言处理后才能表现出来的特征。答案内容特征的量化方式如下:

(1)问题答案主题相似度。高质量答案与对应问题应属于同一个主题[26],使用 LDA 主题模型分别计算答案及其对应问题的主题,再对两者的主题相似性进行分析。

(2)文本多样性。文本多样性由答案文本的用词多样性进行量化得到[27],文本中词汇平均出现的次数越少,答案用词多样性越强。文本多样性的计算方式如公式(1),其中 T_i 为每个词汇在答案中出现的次数。

$$D_{多样性} = \frac{\sum_i^n T_i}{n} \qquad 公式(1)$$

(3)答案信息熵。从信息传播的角度来看,信息熵可以表示信息的价值。高质量答案是具有价值的信息,因此,信息熵在一定程度上可以反映答案的质量[26]。答案信息熵可通过公式(2)计算,其中 P_i 为每个字母在信息中出现的概率。

$$H_{信息熵} = -\sum_i^n P_i * \log_2 P_i \qquad 公式(2)$$

(4)答案情感态度和答案主观性。高质量答案比一般答案更具有情感倾向[16],回答者的态度在一定程度上决定对答案的认可度[26],因此情感态度和主客观性对答案质量存在影响。使用 TextBlob 包(处理文本的 python 库)对答案文本进行情感分析以及主观性判断,答案情感态度及主观性由程序输出的数值结果进行量化得到。

3.1.3 答案其他特征

问答服务中的提问是具有时效性的,提问者通常希望在短时间内得到解答,可见答案的价值与其回答时间负相关。笔者引入"答案回答顺序"作为反映答案时效属性的指标,将所有答案按照回答时间进行升序排列,"答案回答顺序"为当前答案在时间序列中的位次。答案的认可度(推荐量)体现了答案的价值,答案推荐量越高,则该答案是高质量答案的可能性越大。笔者引入"答案推荐量"作为反映其他用户认可度的指标,其是平台自有指标,可由爬虫在 RG 网页中获取。由于"答案回答顺序"和"答案推荐量"无法归入其他维度,故将这两个指标一并归为答案其他特征。

3.1.4 回答者特征

回答者特征关注 RG 平台用户在问答交互中自身对答案质量所造成的影响,通常情况下意见领袖或专业程度更高的专家用户所提供答案的质量更高并能够获得更多用户的支持[15,17]。为了提取出能够反映回答者影响力的特征,笔者结合 RG 平台上可获取到的用户指标,参照意见领袖识别模型 ENIA[28],以 RG-score、RI 值、引文量、个人获赞量、阅读量、项目数等作为回答者专业性、活跃度、影响力等方面的评价指标,相关指标可在回答者个人主页信息中直接获取并量化分析。

3.2 模型选择

笔者选取 ID3 决策树(Iterative Dichotomiser 3, ID3)、随机森林(Random Forest, RF)、支持向量机(Support Vector Machine, SVM)和 BP 神经网络(BP Neural Networks, BP)等答案质量自动化评价任务中的常见模型作为实现学术社交网络平台问答质量智能评价的基础待构建模型。其中,ID3 作为经典的分类算法,虽性能表现一般,但运行速度快、模型可解释性强,满足基线预测模型的要求,因此笔者选择 ID3 作为基线预测模型。当分类模型的性能表现优于 ID3 时,则视为该模型有价值,能够参与组合模型预测。同时,考虑到 RF 适用于数据维度相对较低(几十维)且对准确性有较高要求的情况;而 SVM 在解决小样本下的机器学习问题时有着不错的表现,故选择 RF 和 SVM 作为分类预测模型,符合本研究问答数据维度低(23 维)、数据量小的特点。此外,BP 是目前最常见的深度学习算法之一,拥有强大的特征拟合能力[29],适用于特征之间存在内在联系的情况。由于笔者在构建答案质量评价体系时,为避免数据降维导致重要信息缺失,故尽可能地保留已有变量,使得评价体系存在冗余,评价指标之间存在一定的相关性,因此选用 BP 作为分类模型,符合本研究问答数据特征间存在内在联系的特点。

4 学术社交网络问答质量智能评价实证分析

4.1 数据准备

4.1.1 数据获取与预处理

笔者获取 RG 问答服务中 AI 主题下"Can technology replace a teacher?"问题截至 2020 年 7 月 6 日之前所有的答案及回答者的相关数据，共计 3 873 条。考虑到人工标注的成本问题，笔者采用系统抽样的方式，即从第一条答案开始抽样，以 200 条为抽样间隔距离，最终筛选出 2 000 条答案作为原始数据。对 2 000 条原始数据进行数据清洗，剔除数据中的缺失值、异常值以及乱码数据等，最终得到可用于构建模型的数据 1 670条，并结合 python 自动化程序对各指标进行量化。

4.1.2 数据标注

依据已有问答社区答案质量影响因素研究[30-33]，从实用性、相关性、完整性、可读性及说服力 5 个维度对答案进行评价并完成质量标注。其中，实用性衡量答案是否能帮助解决问题，相关性衡量答案是否和问题相关，完整性衡量答案是否完整、全面地解答问题，可读性衡量答案是否通俗易懂，说服力衡量答案是否可信。

在具体标注过程中，首先邀请具有 RG 问答服务使用经验的 4 名用户作为标注人；然后，对标注人进行培训，使其形成一致的评价标准；最后，采用人工标注的方式对答案质量进行二类标注（1 为高质量，0 为低质量）标注完成后得到带标签的数据集。完成标注的数据集中含有高质量样本 251 条，低质量样本 1 419 条，两者之比接近 1∶6。

4.2 模型的评估与比较

本研究使用 python3.70 编程工具以及 keras、sklearn、imblearn 等 python 库来实现 RG 问答服务答案质量自动化评价模型的搭建。首先，将准备好的数据集自动划分为训练集和测试集，设置划分比例为训练集 80%、测试集 20%；其次，使用训练集分别训练 ID3、RF、SVM、BP 共 4 种算法；最后，在调整参数优化模型后，利用模型在测试集上的查准率（Precision）、查全率（Recall）以及 F1-score 来比较各类算法的性能。其中，查准率代表预测结果中分类正确的正类样本数量占预测结果中所有正类样本数量的比例；查全率表示预测结果中判断正确的正类样本数量占所有真实正类样本数量的比例。查准率和查全率是比较常见的性能评价指标，两个指标的数值与分类器的性能成正比，数值越高模型的效果越好。F1-score 是查准率和查全率的调和平均数，该指标综合考虑两个数值的评估标准，可以对分类器的效果进行更全面的评估[34]。4 种分类模型的性能表现如表 2 所示：

表 2 4 种分类模型预测性能评价（优化前）

模型	0				1			
	Precision	Recall	F1	Support	Precision	Recall	F1	Support
ID3	0.90	0.92	0.91	284	0.48	0.42	0.45	50
RF	0.90	0.92	0.91	284	0.47	0.42	0.44	50
SVM	0.87	0.98	0.92	284	0.59	0.20	0.30	50
BP	0.90	0.95	0.93	284	0.61	0.40	0.48	50

对比上述结果可以发现，4 种模型在对 RG 问答服务的答案质量进行分类预测时，皆出现分类结果向多数类偏移的不平衡分类问题（即分类器趋向于提供一个严重不平衡的精确度，多数类的精确度高，而少数类的精确度却十分低），从而导致预测失败，无法得到准确的结果，模型不具备检测出高质量的答案的能力（查准率、查全率和 F1-score 均偏低）。

这种问题对于学术社交网络问答质量智能评价危害极大，因为问答质量智能评价的最终目的是实现高质量答案的筛选和展示，所以对高质量答案进行正确分类尤为重要。通常而言，针对正负样本数量差别悬殊的情况，通过欠采样或过采样的方法调整数据的分布可以使这种情况得到改善。其中，过采样通过一定的方法使少数类的样本数量增多至和多数种类差不多的个数，从而使得两类样本的数量之比处于相对平衡的状态[35]，其适用于数据量较小且标签样本不平衡的情况，与本文情况相符。为了使模型能够具备筛选高质量答案的能力，笔者将针对 RG 平台答案数据不平衡的特点，从数据层面采用过采样技术改变"标签样本不平衡"的现象，再对高质量答案进行预测。

4.3 基于 SMOTE 算法的模型优化

本研究采用过采样技术 SMOTE 算法（Synthetic Minority Oversampling Technique）[36]来解决标签样本的非平衡问题。SMOTE 算法是基于随机过采样算法的一

种改进方案,是目前处理非平衡数据的常用手段,受到学术界和工业界的一致认同[34]。SMOTE 算法模拟生成新样本的过程采用了 KNN 技术(K-Nearest Neighbor),在生成样本时,不再是对原始样本进行简单地随机复制,因此,生成的新样本更具有代表性。采用 SMOTE 过采样前后产生的训练集和测试集对比如表 3 所示:

表 3 训练样本与测试样本划分

样本数据	原始数据		SMOTE 平衡后	
	高质量答案	低质量答案	高质量答案	低质量答案
总样本	251	1 419	1 419	1 419
训练样本	201	1 135	993	993
测试样本	50	284	426	426

将经 SMOTE 算法平衡后的数据集进行重新划分,利用训练集再次训练 4 种分类模型,并对模型效果进行分析,见表 4。

对比上述结果可以发现,在使用 SMOTE 算法对模型性能进行优化后,4 种模型对高质量答案分类的查准率、查全率都有较大提升,能够满足高质量答案筛选的需求。具体而言,ID3 决策树的性能表现相比于其他几类算法较差;随机森林和 BP 神经网络的性能较好且较为接近;SVM 模型的表现最佳,其性能明显优于其他分类算法,有着最高的查准率、查全率和 F1-score。

表 4 4 种分类模型预测性能评价(优化后)

模型	0				1			
	Precision	Recall	F1	Support	Precision	Recall	F1	Support
ID3	0.88	0.84	0.86	426	0.85	0.89	0.87	426
RF	0.92	0.88	0.90	426	0.89	0.92	0.91	426
SVM	0.95	0.92	0.93	426	0.92	0.95	0.94	426
BP	0.91	0.90	0.90	426	0.90	0.91	0.91	426

4.4 组合模型预测

为进一步发挥不同模型的优势,研究者们通常会使用一定的方法将模型组合起来进行预测。这种组合预测的思想源于 1969 年,由 J. M. Bates 和 C. W. J. Granger 两位学者提出,并普遍应用于机器学习领域。由于随机森林、SVM 和 BP 神经网络 3 个模型性能均优于 ID3 决策树,故全部参与组合预测,组合方式包括两两组合(RF + BP、SVM + BP、SVM + RF)和全模型组合(SVM + RF + BP)。

组合预测是基于“动态加权”的方式实现的,在两两组合情况下,设模型 1 的预测值为 M_1,模型 2 的预测值为 M_2,并分别给两个模型赋予动态权值 w_1 和 w_2。此时,组合模型的预测值 Y 由公式(3)计算得到,其中 $w_1 + w_2 = 1$,w_1 取[0.1, 0.9]递增的 9 个数值。

$$Y = w_1 M_1 + w_2 M_2 \quad \text{公式(3)}$$

同理,全模型组合情况下组合模型的预测值则由 3 个模型的预测结果动态加权得到。随着加权比例的改变,同一组合模型也能得到不同的预测结果,当模型预测效果最好时,得到最优加权比例,4 种组合模型的最优加权结果如表 5 所示:

表 5 组合模型性能评价

组合模型	0				1			
	Precision	Recall	F1	Support	Precision	Recall	F1	Support
0.6 RF + 0.4 BP	0.94	0.92	0.93	426	0.92	0.95	0.93	426
0.7 SVM +0.3 BP	0.95	0.93	0.94	426	0.93	0.96	0.94	426
0.5 SVM +0.5 RF	0.95	0.93	0.94	426	0.94	0.95	0.94	426
0.4 SVM +0.2 RF +0.4 BP	0.96	0.94	0.95	426	0.94	0.96	0.95	426

结果显示,将多个模型进行组合预测后,预测效果较单一模型预测有明显提升。而当 SVM、随机森林、BP 神经网络按照 4:2:4 的比例组合时,模型的预测效果要优于其他两两组合模型,F1-score 达到 96%,模型泛化能力最强。由此可见,借助组合思想构建组合式答案质量预测模型在实现高质量答案筛选时具备有效性,可将组合模型应用于学术社交网络问答质量的智能评价之中。

5 基于问答质量智能评价的问答服务优化

结合前文实证分析结果,在 RG 问答服务功能中提出问答质量智能评价系统架构方案如图 2 所示,以实现学术社交网络平台问答服务的优化。该系统主要包括 3 个部分:问答数据的收集与处理、答案质量评价

以及高质量答案展示。其中,前两个部分是实现答案质量自动化评价的基础,能够对问答服务中的优质回答进行挖掘,最后一个部分聚焦高质量答案展示功能,是用户体验改善的直接推动力。

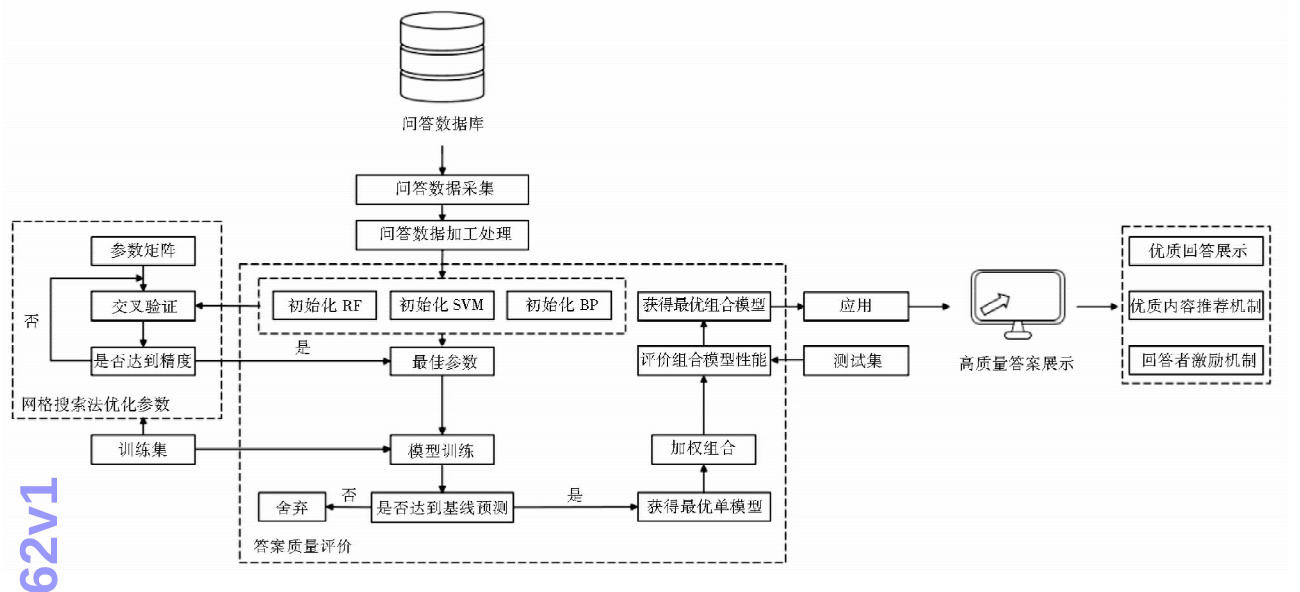


图 2　RG 问答质量智能评价系统架构

5.1　问答数据的收集与处理

系统与 RG 平台的问答数据库进行连接,自动获取答案文本、回答时间、回答者信息等问答数据;将散乱无序的问答数据进行清洗、转换(自动化提取出问答数据中包含的特征)等一系列加工处理过程,形成可供质量评价的原始数据。

5.2　答案质量评价

使用前文证实性能最优的组合式模型作为 RG 平台问答服务答案质量自动化评价模型;将问答数据库中导出的问答数据输入到组合预测模型中,通过计算机程序实现答案质量的自动化预测及标注,完成答案质量评价。

5.3　高质量答案展示

高质量答案展示又细分为 3 个子部分,包括"优质回答展示""优质内容推送机制"和"回答者激励机制",该部分将围绕着问答服务两端的主要参与者(信息接受端即潜水者、提问者;信息发送端即回答者)用户体验的改善展开。

以上系统架构方案同样可为其他学术社交网络平台问答服务版块的优化设计提供参考借鉴。尤其是目前国内的学术社交网络平台尚处于发展初期,平台的问答服务功能仍有待进一步的开发和完善,通过搭建问答质量智能评价系统,能够帮助平台在信息质量问题显现初期对其进行有效的把控,避免负面影响的持续扩大,给用户造成困扰;同时,该系统能够挖掘平台中的优质内容以提升问答服务信息质量水平,促进平台更好地发展。具体而言,学术社交网络平台在实现问答质量智能评价系统的基础上,可在每一个问题的所属页面创建"优质回答"板块,将系统自动筛选出的高质量答案进行展示,使得处在信息接受端的潜水者和提问者能在浏览问题的第一时间内接触到潜在的有用回答,节省用户搜寻回答的时间成本,提高用户的浏览体验。而且人的内心深处都渴望获得重视与欣赏[37],用户的优质回答被公开展示可视为其在学术社交网络平台中的贡献与分享得到肯定,将激发回答者的创作热情,提高他们在学术问答服务中的参与度。同时,在挖掘出潜在优质内容的基础之上,可进一步结合用户专业背景与研究兴趣,建立"优质内容推荐机制",定期向用户推送其可能感兴趣的问题并附上优质回答,促进优质内容的快速传播。此外,考虑到回答者在回答问题后收获的关注或认同将给回答者带来愉悦性,并正向影响回答者回答问题的意愿[38]。因此在回答者的答案被系统认定为优质回答时,以社区积分、荣誉勋章等作为"优质回答激励机制"可激发回答者的认同感和成就感,并促进回答者的持续参与。

6　结论与不足

笔者利用机器学习方法对答案质量进行自动化评价,推动问答质量智能评价的实现,进而促进学术社交网络平台问答信息质量的提升和服务的优化。实证研究结果揭示了从答案结构化特征、答案内容特征、答案其他特征、回答者特征 4 个角度筛选学术社交网络问

答服务中的优质内容(高质量答案)的可行性和合理性,而通过采用机器学习方法,结合 SMOTE 等数据增强技术,并基于组合思想构建组合式答案质量预测模型,则可有效实现高质量答案的筛选。此外,笔者还提出了问答质量智能评价系统设计与实现的整体思路,为 RG 以及其他学术社交网络平台优化其问答服务提供了参考与借鉴。优化方案通过向用户展示优质内容,满足用户对于信息质量的期望;同时,通过展示高质量答案为回答者建立社区认同感和成就感,激励其问答服务的持续参与;最终,通过双向提升问答服务两端用户的粘性,实现学术问答服务的良性循环。

本文研究工作还存在一定局限性。一方面,笔者仅围绕 AI 主题的问答数据,验证答案质量评价体系和自动化评价模型的有效性和合理性,选取的话题内容相对单一,后续研究可在更广范围主题中进行模型的进一步验证分析。另一方面,笔者依据 RG 平台的问答数据展开研究,相关结论和服务优化方案在其他学术社交网络平台或问答社区中的应用效果有待进一步验证。此外,本文研究旨在从学术社交网络平台方揭示问答服务优化路径,后续研究还可进一步从回答者的视角对答案质量进行更加深入的探究。

参考文献:

[1] 李蕾. 学术型社会化问答平台上答案质量评估研究[D]. 南京:南京理工大学,2018.
[2] SHAH C, OH J S, OH S. Exploring characteristics and effects of user participation in online social Q&A sites[J/OL]. First monday, 2008,13(9). [2020-09-25]. https://firstmonday.org/article/view/2182/2028.
[3] 张宁, 袁勤俭. 学术社交网络信息质量的治理和提升[J]. 图书情报工作,2019,63(23):79-86.
[4] 姜雯, 许鑫. 在线问答社区信息质量评价研究综述[J]. 现代图书情报技术,2014(6):41-50.
[5] YAO Y, TONG H, XIE T, et al. Detecting high-quality posts in community question answering sites[J]. Information sciences, 2015,302(1):70-82.
[6] WANG R Y, STOREY V C, FIRTH C P. A framework for analysis of data quality research [J]. IEEE transactions on knowledge & data engineering,1995,7(4):623-640.
[7] TENOPIR C, LEVINE K, ALLARD S, et al. Trustworthiness and authority of scholarly information in a digital age: results of an international questionnaire[J]. Journal of the association for information science & technology,2016,67(10):2344-2361.
[8] WANG R Y, STRONG D M. Beyond accuracy: what data quality means to data consumers[J]. Journal of management information systems,1996,12(4):5-33.
[9] American Public Health Association. Criteria for assessing the quality of health information on the Internet. [J]. American journal of public health,2001,91(3):513-514.
[10] 孙晓宁, 赵宇翔, 朱庆华. 基于SQA系统的社会化搜索答案质量评价指标构建[J]. 中国图书馆学报,2015,41(4):65-82.
[11] DAISUKE I N K, SAKAI O T. What makes a good answer in community question answering? An analysis of assessors' criteria[EB/OL]. [2021-01-05]. https://www.researchgate.net/publication/228449185.
[12] 吴雅威, 张向先, 陶兴, 等. 基于用户感知的学术问答社区答案质量评价指标构建[J]. 情报科学,2020,38(10):141-147.
[13] 张煜轩. 基于外部线索的社会化问答平台答案信息质量感知研究[D]. 武汉:华中师范大学,2016.
[14] CAI Y Z, CHAKRAVARTHY S. Predicting answer quality in Q/A social networks: using temporal features[R]. Arlington: University of Texas at Arlington, 2011.
[15] 孔维泽, 刘奕群, 张敏, 等. 问答社区中回答质量的评价方法研究[J]. 中文信息学报,2011,25(1):3-8.
[16] 姜雯, 许鑫, 武高峰. 附加情感特征的在线问答社区信息质量自动化评价[J]. 图书情报工作,2015,59 (4):100-105.
[17] 郭顺利, 张向先, 陶兴, 等. 社会化问答社区用户生成答案质量自动化评价研究——以"知乎"为例[J]. 图书情报工作,2019,63(11):118-130.
[18] LI L, HE D, JENG W, et al. Answer quality characteristics and prediction on an academic Q&A site: a case study on ResearchGate [C]// 24th international conference on World Wide Web. Florence: ACM, 2015:1453-1458.
[19] LE L T, SHAH C, CHOI E. Assessing the quality of answers autonomously in community question-answering [J]. International journal on digital libraries,2019,20(4):351-367.
[20] VEKARIYA D V, LIMBASIYA N R. A novel approach for semantic similarity measurement for high quality answer selection in question answering using deep learning methods[C]// 6th international conference on advanced computing and communication systems (ICACCS). Coimbatore: IEEE, 2020:518-522.
[21] 贺勋. 在线中文问答社区答案质量预测研究[D]. 济南:齐鲁工业大学,2020.
[22] GOODWIN S, JENG W, HE D. Changing communication on ResearchGate through interface updates [EB/OL]. [2021-01-05]. https://www.researchgate.net/publication/273664849.
[23] LI L, ZHANG C, HE D. Factors influencing the importance of criteria for judging answer quality on academic social Q&A platforms [J]. Aslib journal of information management,2020,72(6):887-907.
[24] LI L, ZHANG C, HE D, et al. Researchers' judgment criteria of high-quality answers on academic social Q&A platforms[J]. Online information review,2020,44(3):603-623.
[25] 任平平. ResearchGate 实现学术社交网络国际化[J]. 国际人才交流,2020(5):52-53.
[26] 王伟, 冀宇强, 王洪伟, 等. 中文问答社区答案质量的评价研

究:以知乎为例[J]. 图书情报工作,2017,61 (22):36 – 44.

[27] 李展，巢文涵，陈小明，等. 中文社区问答中问题答案质量评价和预测[J]. 计算机科学,2011,38(6):230 – 236.

[28] LI Y, MA S, ZHANG Y, et al. An improved mix framework for opinion leader identification in online learning communities[J]. Knowledge-based systems,2013,43(2):43 – 51.

[29] 刘永恒. 基于神经网络和时间序列的汽车销量预测研究[D]. 南昌:南昌大学,2019.

[30] ZHU Z M, BERNHARD D, GUREVYCH I. A multi-dimensional model for assessing the quality of answers in social Q&A sites[EB/OL]. [2020 – 09 – 30]. https://tuprints. ulb. tu-darmstadt. de/1940/.

[31] 贾佳，宋恩梅，苏环. 社会化问答平台的答案质量评估——以“知乎”?“百度知道”为例[J]. 信息资源管理学报,2013,3(2):19 – 28.

[32] 刘乙蓉，刘芸. 问答平台中的答案聚合及其优化:以 Quora 为例[J]. 图书馆学研究,2017(6):48 – 56,13.

[33] 袁红，张莹. 问答社区中询问回答的质量评价——基于百度知道与知乎的比较研究[J]. 数字图书馆论坛,2014(9):43 – 49.

[34] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.

[35] 李宵宵. 随机森林方法在个人信用风险分析中的应用[D]. 昆明:云南大学,2019.

[36] 周琪. 类别不平衡数据的个人信用风险评估算法研究[D]. 保定:河北大学,2020.

[37] MCLAUGHLIN G. SMOG grading-a new readability formula[J]. Journal of reading,1969,12(8):639 – 646.

[38] 张海涛，孙彤，张鑫蕊，等. 社会化问答社区用户角色转变的动力机理研究[J]. 现代情报,2020,40(9): 32 – 41.

作者贡献说明:

严炜炜:提出研究思路,确定选题,修订论文;

黄为:提出研究框架,数据分析,撰写论文;

温馨:负责数据获取和程序编写。

Intelligent Quality Evaluation and Service Optimization of Q&A in Academic Social Networking Site

Yan Weiwei　Huang Wei　Wen Xin

School of Information Management, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] The Q&A service provided by academic social networking site has become an important way for scholars to access academic information quickly and solve academic problems. It is of great significance for the dissemination of high-quality content in academic social networking site to implement the intelligent evaluation of Q&A quality and the service optimization based on machine learning. [**Method/process**] This paper took ResearchGate as the research object, constructed an answer quality evaluation system based on four dimensions of structural features, content features, respondent characteristics and other characteristics of answers, and then used machine learning methods and data augmentation technology to perform the automatic answer quality classification prediction. [**Result/conclusion**] The results show that SMOTE algorithm is effective in dealing with unbalanced samples; In the first mock exam, support vector machine (SVM) achieves excellent classification performance; The combined model can further improve the prediction accuracy, and the combined model based on random forest, SVM and BP neural network has the best classification performance. On this basis, the academic social network Q&A service can be optimized by building the intelligent quality evaluation system.

Keywords: answer quality evaluation　Q&A service　academic social networking site　machine learning